

Application Challenge — Written Submission

Machine Learning Research Engineer @ WindBorne Systems

Neil Dhir, Ph.D. • San Francisco, CA • 15 September 2026

neildhir.github.io • github.com/neildhir • Google Scholar

1. Which of our open ML Research Engineer roles are you interested in, in order?

In order: Applied Research; Model Evaluation; General. Applied, for me, makes sense as I enjoy taking underspecified/ambiguous problems from a first conversation to validated results — see e.g. my DARPA PI role, nuclear fusion or my start-up in Sweden. Model Evaluation, definitely close second as much of my current research is on calibration and uncertainty quantification.

2. Describe an ML idea you were initially excited about but later decided was wrong or unimportant. What changed your mind?

Late in my PhD at Oxford I got terribly excited about infinite-dimensional priors for state-space models (learning the number of latent states from data). Mathematically elegant; practically useless. The test: labelling time-series tens of thousands of hours long. My sampling-based inference couldn't cope; a basic off-the-shelf model did the job perfectly. Heart-breaking but a useful lesson.

3. Describe a time when you embarked on a sidequest / took initiative outside your core job responsibilities. What value did you add to your organization? What drove you to action?

My Focused Energy role doesn't require me to host interns, but I did this summer: fusion is hard and short of qualified people, so you have to engage the next research generation early and get them excited enough to stay. Both were soon writing genuinely useful code; one built a Kalman smoother for a tracking problem that let us buy fewer measuring instruments, saving thousands of dollars.

4. How could better accuracy fail to translate into more useful forecasts for a weather model?

Better point accuracy alone doesn't change decisions. If a pipeline doesn't propagate measurement uncertainty from sensors through to the forecast, you can't attach a calibrated confidence to the number, and users act on confidence against a threshold, not on a slightly better point estimate. The useful claim is 'this accuracy gain cut forecast variance by x%'; without UQ you can never make it.

5. Describe a time you changed how you use LLMs in your work — switching models, tools, or workflows. What did you observe that prompted the change?

I use Claude Code extensively. A few months ago I noticed my agent interactions were strictly sequential when many of the tasks were independent, so I moved to running parallel agents, each allowed to spawn its own sub-agents. Not an 'aha' (or 'Eureka!') moment — more of an 'oh' revelation.

6. Your fav ML Twitter account? (optional)

I regrettably do not have Twitter.